N72-31616

NASA TECHNICAL NOTE

NASA TN D-6935

NASA TN D-6935

CASE FILE COPY

STATISTICAL SUMMARY AND TREND EVALUATION OF AIR QUALITY DATA FOR CLEVELAND, OHIO, IN 1967 TO 1971: TOTAL SUSPENDED PARTICULATE, NITROGEN DIOXIDE, AND SULFUR DIOXIDE

by Harold E. Neustadter, Steven M. Sidik, and John C. Burr, Jr.

Lewis Research Center
Cleveland, Ohio 44135

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • SEPTEMBER 1972

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TN D-6935		

4. Title and Subtitle	5. Report Date
STATISTICAL SUMMARY AND TREND EVALUATION OF AIR QUALITY DATA FOR CLEVELAND, OHIO, IN 1967 TO 1971: TOTAL SUSPENDED PARTICULATE, NITROGEN DIOXIDE, AND SULFUR DIOXIDE	September 1972
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Harold E. Neustadter and Steven M. Sidik, Lewis Research Center; and John C. Burr, Jr., Air Pollution Control Division, Cleveland, Ohio	E-7006

9. Performing Organization Name and Address	10. Work Unit No.
Lewis Research Center National Aeronautics and Space Administration Cleveland, Ohio 44135	770-18
	11. Contract or Grant No.

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
National Aeronautics and Space Administration Washington, D.C. 20546	Technical Note
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Air-quality data for Cleveland, Ohio, for the period of 1967 to 1971 have been collated and subjected to statistical analysis. Total suspended particulate is clearly lognormally distributed, while sulfur dioxide and nitrogen dioxide are reasonably approximated by lognormal distributions. Only sulfur dioxide, in some residential neighborhoods, meets Ohio air-quality standards. Air quality has definitely improved in the industrial valley, while in the rest of the city, only sulfur dioxide has shown consistent improvement. A pollution index has been introduced which displays directly the degree to which the environmental air conforms to mandated standards. As such, it is a useful tool in air-quality monitoring programs.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Total suspended particulate; Sulfur dioxide; Nitrogen dioxide; Air pollution; Pollution index; Cleveland, Ohio	Unclassified - unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	22	\$3.00

* For sale by the National Technical Information Service, Springfield, Virginia 22151

STATISTICAL SUMMARY AND TREND EVALUATION OF AIR QUALITY DATA FOR CLEVELAND, OHIO, IN 1967 TO 1971: TOTAL SUSPENDED PARTICULATE, NITROGEN DIOXIDE, AND SULFUR DIOXIDE

by Harold E. Neustadter, Steven M. Sidik, and John C. Burr, Jr.*

Lewis Research Center

SUMMARY

Air-quality data (total suspended particulate, nitrogen dioxide, and sulfur dioxide) for Cleveland, Ohio, for the period of 1967 to 1971 have been collated and subjected to statistical analysis. Total suspended particulate is clearly lognormally distributed, while sulfur dioxide and nitrogen dioxide are reasonably approximated by lognormal distributions. The air-quality standards for the State of Ohio are met only sporadically by sulfur dioxide in isolated residential neighborhoods. Nowhere in Cleveland are the standards for total suspended particulate or nitrogen dioxide met. Definite improvement in air quality has taken place in the industrial valley, while in the rest of the city, only sulfur dioxide has shown consistent improvement.

A pollution index has been introduced which directly displays information regarding the degree to which the environmental air conforms to the mandated standards. As such, it is a useful tool in air-quality monitoring programs.

INTRODUCTION

This report presents the results of various statistical analyses of data obtained by the Air Pollution Control Division (APCD) of Cleveland, Ohio. It contains a tabulation of averages, statistics relevant to lognormal distributions, and goodness-of-fit statistics. In addition, a pollution-level index is introduced which relates the measured pollution levels over a year to the existing air-quality standards.

*Air Pollution Control Division, Cleveland, Ohio.

The air-sampling program of APCD is currently in its sixth year. Twenty-four-hour samplings have been made of total suspended particulate (TSP) since January 1967, and of nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) since January 1968. The sampling methods used are high-volume air sampling, Jacobs-Hocheiser, and West-Gaeke sulfuric acid, respectively. The geographic deployment of sampling sites is shown in figure 1. The meandering heavy line in the center of the city is the Cuyahoga River, about which is centered most of the region's heavy industry.

At present, there are 21 stations monitoring the air. Fifteen of these stations monitor all three pollutants, while the remaining six (stations O to T in fig. 1) measure TSP only. Seventeen of these sites have been in operation for more than 5 years. Stations B, D, K, and N have undergone relocation since their initial installation. However, because of the proximity of their present sites to their former sites, we have assumed that essentially the same environment has been measured throughout the 5-year

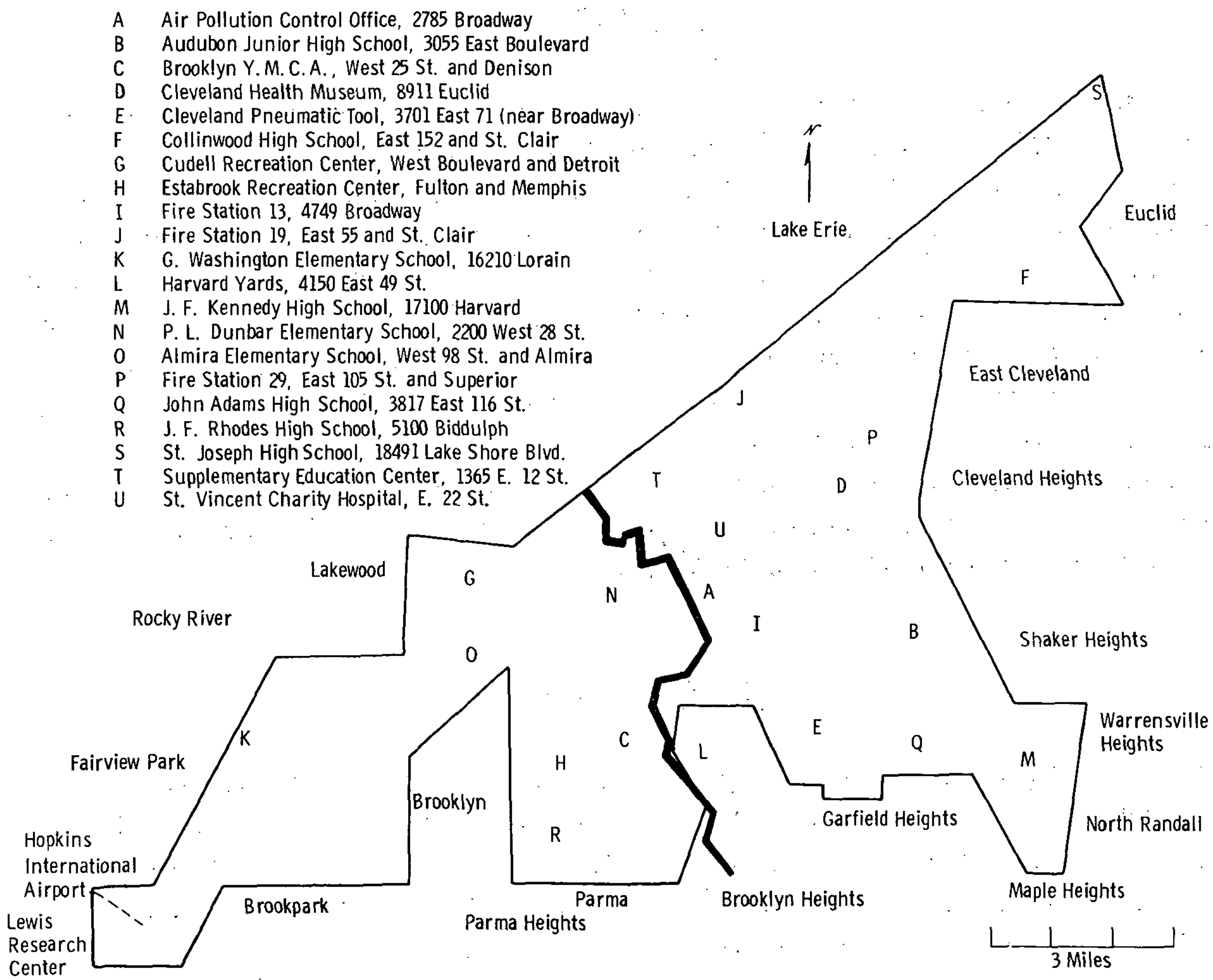


Figure 1. - Air pollution monitoring sites for Cleveland, Ohio. The heavy line down the center is the Cuyahoga River. The municipal boundaries have been straightened somewhat but are accurate in their essential features.

period. Currently, the air is sampled every third day, although the sampling frequency has varied over the 5 years and has been as low as once a week. Some of these data have been presented elsewhere in a more preliminary manner (ref. 1). The data analysis reported herein was performed by the Environmental Research Office of the NASA Lewis Research Center (LeRC) as part of the preliminary phase of a joint APCD-LeRC program to study trace elements and compounds in airborne particulate matter.

CLEVELAND AEROMETRIC DATA

Pertinent results are presented in tables I, II, and III for TSP, NO_2, and SO_2, respectively. In each table, the first column gives an alphabetic designation of the monitoring site corresponding to the code shown in figure 1. The second column lists the various parameters of interest for each of the pollutants. These parameters are (1) number of days observed (readings); (2) geometric (TSP) or arithmetic (SO_2 and NO_2) averages; (3) standard geometric deviation; (4) estimated value of the second largest pollution level for the year; and (5) an adjusted Kolmogorov-Smirnov goodness-of-fit statistic for lognormality, denoted as $\sqrt{N}\,D$.

Air-quality standards are set nationally by the Environmental Protection Agency (EPA) of the Federal Government (ref. 2) and statewide by the Air Pollution Control Board of the Department of Health (DoH) of the State of Ohio (ref. 3). Whenever these two standards differ, we have chosen to work with the DoH (more stringent) standard, which is listed in the third column. In the remaining five columns are the various statistics for each of the years 1967 to 1971.

Number of Readings

For each pollutant, both EPA and DoH require a minimum of one sampling every sixth day, or an equivalent set of at least 61 random samples per year. Thus, we designate this standard as >60 in the tables. Even though early in the program some stations did not achieve 60 samples per year for each pollutant, we have included the analyses of these data sets in this report. At present, the nominal schedule of APCD calls for monitoring the environmental air every third day. In practice, this procedure generally allows sufficient margin for unanticipated disruptions (e.g., equipment failure) while still exceeding 60 readings per year.

TABLE I. - TOTAL SUSPENDED PARTICULATE DATA SUMMARY FOR 1967 TO 1971

Monitoring station (see fig. 1)	Statistic	Standard	1967	1968	1969	1970	1971
A	Number of readings	>60	19	70	73	76	69
	Geometric average	60	190	242	199	188	183
	Standard geometric deviation		1.4	1.7	1.6	1.6	1.7
	Second highest reading	50		919	[a]711	[a]682	730
	Goodness-of-fit statistic, $\sqrt{N}\,D$			0.53	0.84	0.81	0.73
B	Number of readings	>60	36	64	66	[b]72	63
	Geometric average	60	112	104	94	113	92
	Standard geometric deviation		1.5	1.6	1.4	1.6	1.6
	Second highest reading	150	351	349	226	370	319
	Goodness-of-fit statistic, $\sqrt{N}\,D$		0.76	0.72	0.63	0.48	0.53
C	Number of readings	>60	64	79	72	97	89
	Geometric average	60	124	121	107	124	121
	Standard geometric deviation		1.5	1.6	1.6	1.6	1.7
	Second highest reading	150	343	[a]429	346	420	502
	Goodness-of-fit statistic, $\sqrt{N}\,D$		0.55	0.76	0.50	0.39	0.65
D	Number of readings	>60	44	72	74	[b]62	[c]30
	Geometric average	60	134	126	123	154	163
	Standard geometric deviation		1.5	1.5	1.5	1.6	1.8
	Second highest reading	150	371	390	378	487	
	Goodness-of-fit statistic, $\sqrt{N}\,D$		0.37	0.42	0.50	0.40	
E	Number of readings	>60	61	75	75	93	80
	Geometric average	60	139	147	119	136	120
	Standard geometric deviation		1.4	1.5	1.4	1.5	1.5
	Second highest reading	150	352	[a]410	276	[a]395	[a]328
	Goodness-of-fit statistic, $\sqrt{N}\,D$		0.59	0.83	0.61	0.80	0.80
F	Number of readings	>60	64	75	75	82	74
	Geometric average	60	101	103	88	109	105
	Standard geometric deviation		1.5	1.6	1.6	1.5	1.5
	Second highest reading	150	[a]303	357	297	[a]307	304
	Goodness-of-fit statistic, $\sqrt{N}\,D$		1.0	0.67	0.64	0.87	0.72
G	Number of readings	>60	8	75	73	103	83
	Geometric average	60		99	82	94	91
	Standard geometric deviation			1.6	1.6	1.7	1.6
	Second highest reading	150		317	[a]292	358	337
	Goodness-of-fit statistic, $\sqrt{N}\,D$			0.56	0.79	0.59	0.57
H	Number of readings	>60		65	68	96	70
	Geometric average	60		83	84	94	89
	Standard geometric deviation			1.6	1.6	1.7	1.7
	Second highest reading	150		280	299	384	352
	Goodness-of-fit statistic, $\sqrt{N}\,D$			0.53	0.59	0.48	0.68
I	Number of readings	>60	55	75	75	101	93
	Geometric average	60	210	232	223	225	196
	Standard geometric deviation		1.4	1.5	1.5	1.5	1.6
	Second highest reading	150	[a]543	694	[a]639	701	[a]658
	Goodness-of-fit statistic, $\sqrt{N}\,D$		1.08	0.60	0.97	0.51	0.83
J	Number of readings	>60	63	76	74	103	90
	Geometric average	60	174	161	151	156	163
	Standard geometric deviation		1.5	1.6	1.7	1.6	1.7
	Second highest reading	150	474	[a]538	[a]613	[a]530	645
	Goodness-of-fit statistic, $\sqrt{N}\,D$		0.62	0.78	0.76	0.98	0.73
K	Number of readings	>60	75	80	75	[b]87	78
	Geometric average	60	85	81	73	88	92
	Standard geometric deviation		1.5	1.6	1.6	1.5	1.6
	Second highest reading	150	[a]254	[a]273	246	257	312
	Goodness-of-fit statistic, $\sqrt{N}\,D$		0.96	0.92	0.68	0.68	0.52

[a]The calculation used to obtain this estimate assumed lognormality despite $\sqrt{N}\,D \geq 0.736$.
[b]Sampling site was relocated within same general neighborhood in midyear. It is assumed that for sampling purposes the environmental air was the same at both locations.
[c]Temporarily discontinued because of construction at sampling site.

TABLE I. - Concluded. TOTAL SUSPENDED PARTICULATE DATA SUMMARY FOR 1967 TO 1971

Monitoring station (see fig. 1)	Statistic	Standard	1967	1968	1969	1970	1971
L	Number of readings	>60				37	73
	Geometric average	60				170	212
	Standard geometric deviation					1.5	1.6
	Second highest reading	150				525	637
	Goodness-of-fit statistic, $\sqrt{N}$ D					0.49	0.64
M	Number of readings	>60	60	72	74	89	72
	Geometric average	60	86	82	75	86	82
	Standard geometric deviation		1.5	1.6	1.5	1.6	1.6
	Second highest reading	150	266	281	222	294	284
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.48	0.64	0.60	0.62	0.59
N	Number of readings	>60	48	75	73	[b]75	86
	Geometric average	60	129	158	142	134	138
	Standard geometric deviation		1.8	1.8	1.9	2.4	2.0
	Second highest reading	150	592	784	747	[a]1273	905
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.60	0.57	0.67	0.99	0.71
O	Number of readings	>60	69	75	72	90	76
	Geometric average	60	92	86	79	89	90
	Standard geometric deviation		1.5	1.6	1.6	1.7	1.8
	Second highest reading	150	265	298	[a]270	333	422
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.62	0.39	0.83	0.71	0.55
P	Number of readings	>60	62	74	72	93	74
	Geometric average	60	135	139	127	137	146
	Standard geometric deviation		1.4	1.5	1.6	1.5	1.4
	Second highest reading	150	343	390	407	412	371
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.71	0.40	0.64	0.55	0.60
Q	Number of readings	>60	63	69	70	88	79
	Geometric average	60	105	95	96	106	101
	Standard geometric deviation		1.5	1.5	1.4	1.8	1.4
	Second highest reading	150	310	277	241	[a]495	256
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.62	0.42	0.67	0.97	0.65
R	Number of readings	>60	57	72	65	90	66
	Geometric average	60	81	80	81	89	89
	Standard geometric deviation		1.6	1.7	1.6	1.6	1.7
	Second highest reading	150	265	304	285	309	384
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.44	0.69	0.52	0.49	0.60
S	Number of readings	>60					51
	Geometric average	60					92
	Standard geometric deviation						1.5
	Second highest reading	150					290
	Goodness-of-fit statistic, $\sqrt{N}$ D						0.71
T	Number of readings	>60					41
	Geometric average	60					170
	Standard geometric deviation						2.0
	Second highest reading	150					1014
	Goodness-of-fit statistic, $\sqrt{N}$ D						0.48
U	Number of readings	>60					[d]26
	Geometric average	60					162
	Standard geometric deviation						1.5
	Second highest reading	150					
	Goodness-of-fit statistic, $\sqrt{N}$ D						

[a]The calculation used to obtain this estimate assumed lognormality despite $\sqrt{N}\,D \geq 0.736$.
[b]Sampling site was relocated within same general neighborhood in midyear. It is assumed that for sampling purposes the environmental air was the same at both locations.
[c]Temporarily discontinued because of construction at sampling site.
[d]Sampling was initiated in the latter part of the year.

TABLE II. - NITROGEN DIOXIDE DATA SUMMARY FOR 1968 TO 1971

Monitoring station (see fig. 1)	Statistic	Standard	1968	1969	1970	1971
A	Number of readings	>60	71	73	84	86
	Geometric average	100	211	220	214	202
	Standard geometric deviation		1.4	1.4	1.4	1.5
	Second highest reading		517	470	464	538
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.60	0.57	0.61	0.59
B	Number of readings	>60			9	81
	Geometric average	100				190
	Standard geometric deviation					1.5
	Second highest reading					[a]539
	Goodness-of-fit statistic, $\sqrt{N}$ D					0.77
C	Number of readings	>60	76	75	115	96
	Geometric average	100	177	248	234	255
	Standard geometric deviation		1.5	1.3	1.4	1.6
	Second highest reading		[a]495	[a]454	[a]576	835
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.87	0.88	0.88	0.64
D	Number of readings	>60	55	70	[b]83	[c]47
	Geometric average	100	207	219	217	199
	Standard geometric deviation		2.0	1.3	1.5	1.4
	Second highest reading		[a]1056	424	[a]576	465
	Goodness-of-fit statistic, $\sqrt{N}$ D		1.65	0.70	1.03	0.62
E	Number of readings	>60	69	74	108	96
	Geometric average	100	203	237	217	205
	Standard geometric deviation		1.4	1.3	1.4	1.6
	Second highest reading		497	[a]437	[a]504	[a]686
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.70	0.90	1.39	1.69
F	Number of readings	>60	47	74	96	86
	Geometric average	100	212	197	215	203
	Standard geometric deviation		1.4	1.3	1.3	1.5
	Second highest reading		[a]511	[a]370	444	[a]518
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.78	0.76	0.70	0.93
G	Number of readings	>60	72	72	104	89
	Geometric average	100	201	221	224	203
	Standard geometric deviation		1.5	1.3	1.3	1.5
	Second highest reading		571	[a]432	453	516
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.56	0.91	0.43	0.65
H	Number of readings	>60	66	71	114	78
	Geometric average	100	166	225	213	202
	Standard geometric deviation		1.5	1.3	1.4	1.6
	Second highest reading		[a]471	[a]443	464	[a]633
	Goodness-of-fit statistic, $\sqrt{N}$ D		1.03	0.75	0.70	1.1
I	Number of readings	>60	67	76	111	88
	Geometric average	100	247	253	238	217
	Standard geometric deviation		1.4	1.3	1.3	1.5
	Second highest reading		535	495	[a]495	[a]615
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.45	0.71	1.1	0.93
J	Number of readings	>60		52	113	93
	Geometric average			225	255	240
	Standard geometric deviation			1.4	1.4	1.5
	Second highest reading			488	[a]548	600
	Goodness-of-fit statistic, $\sqrt{N}$ D			0.65	0.82	0.58
K	Number of readings	>60	74	74	[b]104	88
	Geometric average	100	162	192	209	183
	Standard geometric deviation		1.5	1.4	1.4	1.6
	Second highest reading		433	417	[a]486	565
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.53	0.67	0.76	0.67
L	Number of readings	>60			41	80
	Geometric average				220	219
	Standard geometric deviation				1.4	1.5
	Second highest reading				513	572
	Goodness-of-fit statistic, $\sqrt{N}$ D				0.68	0.71
M	Number of readings	>60	55	74	96	73
	Geometric average		157	168	176	159
	Standard geometric deviation		1.4	1.3	1.3	1.6
	Second highest reading		[a]342	335	341	507
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.80	0.60	0.65	0.54
N	Number of readings	>60			39	88
	Geometric average				208	223
	Standard geometric deviation				1.6	1.6
	Second highest reading				647	[a]712
	Goodness-of-fit statistic, $\sqrt{N}$ D				0.65	0.95
U	Number of readings	>60				[d]36
	Geometric average					230
	Standard geometric deviation					1.9
	Second highest reading					[a]1030
	Goodness-of-fit statistic, $\sqrt{N}$ D					1.34

[a]The calculation used to obtain this estimate assumed lognormality despite $\sqrt{N}$ D $\geq$ 0.736.

[b]Sampling site was relocated within same general neighborhood in midyear. It is assumed that for sampling purposes the environmental air was the same at both locations.

[c]Temporarily discontinued because of construction at sampling site.

[d]Sampling was initiated in the latter part of the year.

TABLE III. - SULFUR DIOXIDE DATA SUMMARY FOR 1968 TO 1971

Monitoring station (see fig. 1)	Statistic	Standard	1968	1969	1970	1971
A	Number of readings	>60	71	74	82	88
	Geometric average	60	137	135	116	84
	Standard geometric deviation		2.4	2.0	1.9	2.2
	Second highest reading	260	[a]972	[a]674	[a]518	523
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.75	0.96	0.88	0.66
B	Number of readings	>60			9	86
	Geometric average	60				50
	Standard geometric deviation					2.1
	Second highest reading	260				284
	Goodness-of-fit statistic, $\sqrt{N}$ D					0.70
C	Number of readings	>60	72	76	105	93
	Geometric average	60	95	85	74	67
	Standard geometric deviation		2.4	2.3	2.3	2.4
	Second highest reading	260	644	546	476	485
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.61	0.48	0.54	0.73
D	Number of readings	>60	53	72	[b]79	[c]45
	Geometric average	60	106	103	109	89
	Standard geometric deviation		1.8	1.7	2.0	2.0
	Second highest reading	260	413	278	[a]538	[a]469
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.52	0.47	0.91	0.76
E	Number of readings	>60	71	75	107	94
	Geometric average	60	112	107	96	65
	Standard geometric deviation		1.9	1.6	1.8	2.1
	Second highest reading	260	476	314	[a]397	375
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.68	0.42	0.88	0.71
F	Number of readings	>60	47	75	97	86
	Geometric average	60	84	76	90	59
	Standard geometric deviation		1.9	2.1	1.8	2.3
	Second highest reading	260	[a]364	[a]409	373	[a]401
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.80	1.04	0.68	0.83
G	Number of readings	>60	69	71	105	86
	Geometric average	60	77	58	63	50
	Standard geometric deviation		2.1	2.0	1.9	2.4
	Second highest reading	260	414	294	295	[a]363
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.57	0.70	0.70	0.75
H	Number of readings	>60	62	71	113	72
	Geometric average	60	64	63	66	48
	Standard geometric deviation		2.3	2.3	2.2	2.4
	Second highest reading	260	[a]416	390	408	336
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.85	0.69	0.47	0.72

Monitoring station (see fig. 1)	Statistic	Standard	1968	1969	1970	1971
I	Number of readings	>60	64	77	108	83
	Geometric average	60	129	110	101	67
	Standard geometric deviation		1.8	1.8	1.9	2.1
	Second highest reading	260	[a]522	467	[a]449	[a]358
	Goodness-of-fit statistic, $\sqrt{N}$ D		1.04	0.64	0.87	0.90
J	Number of readings	>60		52	113	93
	Geometric average	60		113	124	79
	Standard geometric deviation			1.9	1.8	2.0
	Second highest reading	260		543	504	[a]410
	Goodness-of-fit statistic, $\sqrt{N}$ D			0.53	0.70	1.23
K	Number of readings	>60	74	75	[b]105	81
	Geometric average	60	53	58	59	49
	Standard geometric deviation		2.5	2.1	1.9	2.4
	Second highest reading	260	399	320	258	[a]359
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.55	0.57	0.64	0.83
L	Number of readings	>60			42	79
	Geometric average	60			157	116
	Standard geometric deviation				1.7	2.6
	Second highest reading	260			569	[a]1013
	Goodness-of-fit statistic, $\sqrt{N}$ D				0.62	0.98
M	Number of readings	>60	53	73	98	58
	Geometric average	60	50	55	58	41
	Standard geometric deviation		1.9	1.9	2.3	2.6
	Second highest reading	260	220	235	309	[a]372
	Goodness-of-fit statistic, $\sqrt{N}$ D		0.72	0.67	0.67	0.74
N	Number of readings	>60			35	81
	Geometric average	60			68	72
	Standard geometric deviation				2.6	2.9
	Second highest reading	260			[a]548	[a]755
	Goodness-of-fit statistic, $\sqrt{N}$ D				0.76	0.90
U	Number of readings	>60				[d]34
	Geometric average	60				114
	Standard geometric deviation					2.3
	Second highest reading	260				137
	Goodness-of-fit statistic, $\sqrt{N}$ D					0.55

[a]The calculation used to obtain this estimate assumed lognormality despite $\sqrt{N}\,D \geq 0.736$.

[b]Sampling site was relocated within same general neighborhood in midyear. It is assumed that for sampling purposes the environmental air was the same at both locations.

[c]Temporarily discontinued because of construction at sampling site.

[d]Sampling was initiated in the latter part of the year.

Geometric and Arithmetic Averages

The geometric average is used in table I, and the arithmetic average is used in tables II and III. This corresponds to the particular averaging method stipulated by EPA and DoH standards. Calculations were performed whenever the number of readings exceeded 10. The values listed as standards are the DoH primary standards, which correspond to the EPA secondary standards.

Standard Geometric Deviation (SGD)

It has been noted that, irrespective of sampling duration or location, air-sampling data are generally distributed lognormally (ref. 4). When such is actually the case, the entire data set is sufficiently described by its geometric average and SGD. The higher the SGD, the greater the spread between the lower and higher values. As with the averages, SGD was calculated for data sets of more than 10 readings.

Second Largest Value

Both EPA and DoH standards for TSP and SO_2 specify that a certain level of pollution is ". . . not to be exceeded more than one time per year." This implies that for the 365 daily pollution levels per year (366 for leap years), there is no upper bound on the largest single level. However, the next largest value (i. e., the second most polluted day of the year) is required to be at or below the standard. Thus, tables I, II, and III include estimates of the second highest pollution level for each year. As with the averages, the values listed here are the DoH primary standards, which correspond to EPA secondary standards. While NO_2 has only a standard for the annual average, we believe the estimated second largest level for a year is useful information and we have included it in table II.

An approximation to the second largest pollution level estimate, for a year of n days and a sample of N observations, is obtained by the following procedure. (The transformation to the logarithms of the data values is made because the expected values of normal order statistics are well developed in the literature, whereas we are not aware of any comparable development for lognormal distributions.) The logarithms $y_i = \ln(x_i)$ of the pollution levels x_i are computed. According to the assumption of lognormality, these y_i values follow a normal distribution. The sample mean $\bar{y}$ and sample standard deviation s_y of the set of logarithms are computed. From reference 4, the expected value of the second largest observation in a sample of 365 (366 in a leap year) independent values from a normal distribution is 2.63 (to three significant digits)

standard deviations from the mean. This value, along with the average $\bar{y}$ and the standard deviation s_y of the set of logarithms, is used in the following equation to obtain the estimate of the second largest pollution level of the year:

$$y_{2nd} = \bar{y} + 2.63\ s_y \tag{1}$$

The values of x_{2nd} listed in tables I, II, and III are obtained by exponentiation, as

$$x_{2nd} = \exp(y_{2nd}) \tag{2}$$

Because of the decreased precision which occurs when extrapolating to the tail of a distribution and because the sample mean and standard deviation are used, the minimum number of readings for this calculation was increased to 30 as opposed to 10 used for the averages. Implicit in using equation (1) is the assumption of lognormality of the data, which leads us to the final entry in these tables.

Kolmogorov-Smirnov Statistic

The Kolmogorov-Smirnov statistic is a goodness-of-fit statistic which can be applied to any distribution (ref. 5). In testing for a lognormal distribution, it is easier for calculation purposes to take the logarithms of the values and test for goodness-of-fit to a normal distribution. This statistic was originally intended for use when the distribution which the data is suspected of following is completely specified. For the normal distribution, this is equivalent to knowing the mean μ and the standard deviation σ. In this case, the Kolmogorov-Smirnov statistic is denoted D and is calculated as

$$D = \max_{i = 1, N} \left| \Phi\left(\frac{y_i - \mu}{\sigma}\right) - \left(\frac{i}{N}\right) \right| \tag{3}$$

where the function $\Phi(z)$ denotes the cumulative standard normal distribution function.

The statistic D measures the maximum deviation of the observed cumulative distribution function from the theoretical cumulative distribution function. Thus, D is always a value between zero and one. A value of zero would indicate a perfect fit of the sampled data to a lognormal distribution, and larger values indicate an increasing deviation from lognormality.

When the mean and the standard deviation are unknown, it is common to use the estimates $\bar{y}$ and $s_y = \left[\Sigma_i(y_i - \bar{y})^2/(N - 1)\right]^{1/2}$ in place of μ and σ. Lilliefors has studied the use of the Kolmogorov-Smirnov statistic in this situation (ref. 6). Table IV

TABLE IV. - SIGNIFICANCE LEVELS FOR THE KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT STATISTIC

[From ref. 6.]

Significance level, α	0.20	0.15	0.10	0.05	0.01
Statistic, $\sqrt{N}\,D_{\alpha}$	0.736	0.768	0.805	0.886	1.031

of this report presents the significance levels of $\sqrt{N}\,D$ from reference 6 for samples of $N > 30$. Thus, the statistics in tables I, II, and III are presented as $\sqrt{N}\,D$.

It should be recognized that the observed pollution levels are but a sample of levels from some distribution. Thus, even if the distribution of the complete set of pollution levels is indeed lognormal, some of the samples will lead to large values of $\sqrt{N}\,D$. The interpretation of the tabulated significance levels α is that if the distribution is indeed lognormal, then about 100α percent of the samples tested will lead to a value of $\sqrt{N}\,D$ which exceeds $\left(\sqrt{N}\,D\right)_{\alpha}$, whereas about $100(1 - \alpha)$ percent will lead to a value of $\sqrt{N}\,D$ lower than $\left(\sqrt{N}\,D\right)_{\alpha}$. Because subsequent calculations in this report depend heavily on the assumption of lognormality, the value of $\alpha = 0.20$ was chosen. Choosing this large value for α has the drawback of rejecting the assumption of lognormality a substantial proportion of the times that the distribution is lognormal. However, it has the compensating advantage of being more discriminating against distributions which are not lognormal.

LOGNORMALITY

Lognormal Plots

As a graphical means of assessing the goodness-of-fit of the data to a lognormal distribution, we can enter the observed data on lognormal probability graphs. Figures 2 and 3 show two such plots for TSP. The solid line indicates the plot of the cumulative sample distribution of all measurements over the 5-year period. The data points present the separate sample distributions for the 5 years (1967 to 1971). Any steady increase or decrease in the pollutant concentrations would be discernible as a vertical sequence of the data points representing those years. In the two cases shown, there is

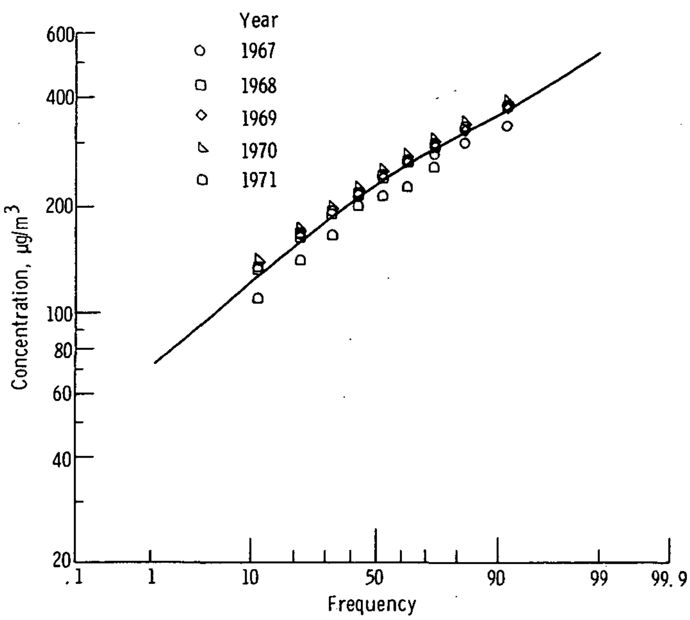


Figure 2. - Lognormal plot of distribution by weight of total suspended particulate (24-hr sampling) for monitoring station I (see fig. 1) downwind of the industrial region.

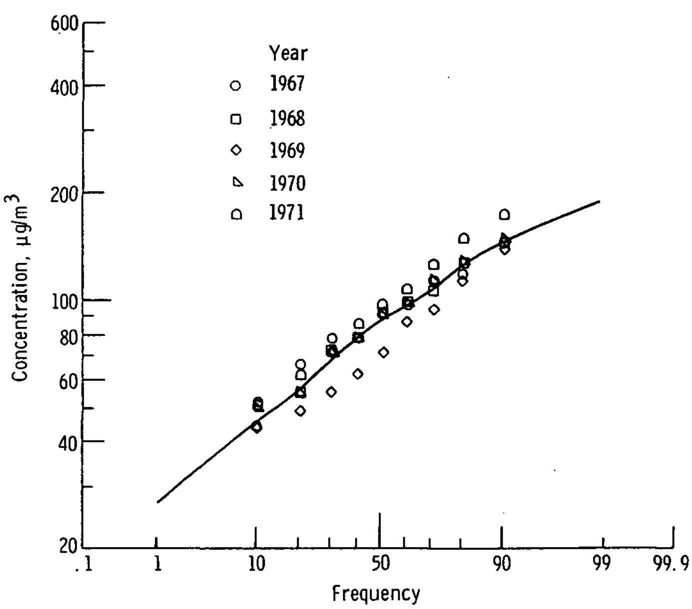


Figure 3. - Lognormal plot of distribution by weight of total suspended particulate (24-hr sampling) for monitoring station K (see fig. 1) upwind of the industrial region.

no overall trend. Figure 2 is for station I in the industrial valley. The overprinting of the data points shows the TSP levels to be fairly uniform at a rather high average level for the 5-year period. Figure 3 represents station K, in a residential neighborhood, predominantly upwind from the industrial region.

A full set of lognormal curves for all 21 stations for the three pollutants is available on microfiche from the authors upon request.

Goodness of Fit

To indicate the decreasing likelihood of lognormality as $\sqrt{N}\,D$ increases, all values calculated on the assumption of lognormality for which the goodness-of-fit statistic is outside the 20-percent confidence level (i.e., the data having $\sqrt{N}\,D > 0.736$) are footnoted in the tables. For a further indication of lognormality, as well as for a check on the consistency of our data, we examined the distribution of sets for which $\sqrt{N}\,D > 0.736$.

Table V summarizes the results of the goodness-of-fit tests in which the $\alpha = 0.20$ significance level was used. The first column lists the station identification. The remaining columns list for each of the pollutants the number of yearly tests which were performed and the number of these tests which rejected the assumption of lognormality. For TSP, there are 85 tests, of which 20 were rejections. This is very close to the expected number of rejections and implies that the distribution of TSP may very safely be considered to be lognormal. For NO_2 and SO_2, however, there are more than twice as many rejections as would be expected, and hence their closeness to a lognormal distribution is somewhat suspect. On the basis of an examination of the lognormal plots of SO_2 and the fact that the SO_2 departure from lognormality, as indicated by $\sqrt{N}\,D$, is not severe, we will proceed on the assumption that the lognormal is still a useful approximation to the distribution of SO_2.

Further examination of table V shows that the lognormality of TSP, SO_2 and NO_2 is most questionable at stations E, F, and I. Benarie (ref. 7) and Mitchell (ref. 8) have each considered the additivity of lognormal distributions. Mitchell has shown that under certain conditions the sum of independent and identically distributed lognormal variates also follows a lognormal distribution. Benarie has considered a more general situation, where the lognormal variates have differing geometric means and standard geometric deviations. His conclusions are that when a large number (>10) of lognormal variates with slightly differing geometric means are superimposed, the resulting distribution is still well approximated by a lognormal distribution. However, when a small number (<10) of lognormal variates with differing means are superimposed, the resulting distribution generally is not a lognormal. Thus, it is possible to conjecture that pollution

TABLE V. - SUMMARY OF RESULTS OF GOODNESS-OF-FIT TESTS

Monitoring station (see fig. 1)	Total suspended particulate		Nitrogen dioxide		Sulfur dioxide	
	Number of tests	Rejected	Number of tests	Rejected	Number of tests	Rejected
A	4	2	4	0	4	3
B	5	0	1	1	1	0
C	5	1	4	3	4	0
D	4	0	4	2	4	2
E	5	3	4	3	4	1
F	5	2	4	3	4	3
G	4	1	4	1	4	1
H	4	0	4	3	4	1
I	5	3	4	2	4	3
J	5	3	3	1	3	1
K	5	2	4	1	4	1
L	2	0	2	0	2	1
M	5	0	4	1	4	1
N	5	1	2	1	2	2
O	5	1				
P	5	0				
Q	5	1				
R	5	0				
S	1	0				
T	1	0				
U			1	1	1	0
Total	85	20	49	23	49	20
Percentage of tests rejected		24		47		41
Expected number of rejections		17		9.8		9.8

levels at stations E, F, and I are dominated by a small number of major sources, whereas the remaining stations reflect the influence of either a single large source or a superposition of many sources.

AIR QUALITY

Among the goals of APCD are monitoring of the environmental air, determination

of its quality, and initiation of action to improve the local air quality, where indicated. There are well established techniques for analysing lognormal plots to extract information pertinent to determining compliance with air-quality standards and/or the existence of long-term trends (ref. 9). However, it is often desirable to have available some single number, or index, which presents as simply as possible a maximum of information. To this end we have developed an index, which we call Polludex, which gages the conformity of the measured environment to the established standards.

Polludex, An Air-Pollution Index

Many indices have been proposed and a number are in use by various agencies (ref. 10). Polludex is a variation of an index proposed by Pikul (ref. 11). The rationale for constructing this modified index is as follows. The standards for TSP and SO_2 specify values for the annual mean which may not be exceeded and also values which may not be exceeded more than once per year. In relation to a lognormal plot of the underlying population, these standard values specify the coordinates of two points on a straight line. If the data obtained during a 1-year period conform to lognormality and conform to the required standards, the plot of the data will closely approximate a straight line falling entirely below (or on) the line segment joining the standard points.

For each of the three pollutants, define

$$r = \frac{\text{Sample average}}{\text{Standard for average}}$$

$$s = \frac{\text{Estimate of second largest level}}{\text{Standard not to be exceeded more than once yearly}}$$

Then Polludex, P (pollutant), is defined for TSP and SO_2 by

$$P(TSP, SO_2) = 50 \times \left[\max(0, r - 1) + \max(0, s - 1)\right]$$

and for NO_2 by

$$P(NO_2) = 100 \times \left[\max(0, r - 1)\right]$$

where $\max(a, b)$ means that the larger of the two values, a or b, is to be used. The geometric average is to be used in calculating r for TSP and the arithmetic average is to be used in calculating r for SO_2 and NO_2. For the estimate of the second largest

level to be used for s we used the approximate value listed in table I for TSP and in table III for SO_2.

With this definition, the same weight is given to the long-term (chronic) effects of pollution as is given to the severe short-term (episode) incident. The standards for these pollutants have presumably been set with regard to maximum acceptable levels for reasons of public health and/or welfare. Thus, we assume that normalization of the estimated mean and second highest values by the standards will, in a sense, put each P on an equal basis with respect to the potential harm caused by excesses. If the air quality is equal to or better than the standards, Polludex = 0. A value of Polludex = 100 can be understood to mean that the air is, in a sense, 100 percent polluted, in that a value of 100 is obtained when the average and the second highest values are each 100 percent higher than their respective permissible levels. Of course, Polludex = 100 would also result from a continuum of other combinations, as, for example, when the second highest value is three times its standard, provided the average was at or below its standard. Figure 4 graphically illustrates several of these possibilities. Figure 4(a) shows three possible examples which have P = 0. Figure 4(b) shows a line having P = 100, where both the mean and second largest standards are exceeded. Figure 4(c) shows a line where again P = 100, but the standard for the mean has been met. Finally, figure 4(d) shows a line with P = 50, where the standard for the mean is not met but the other standard is.

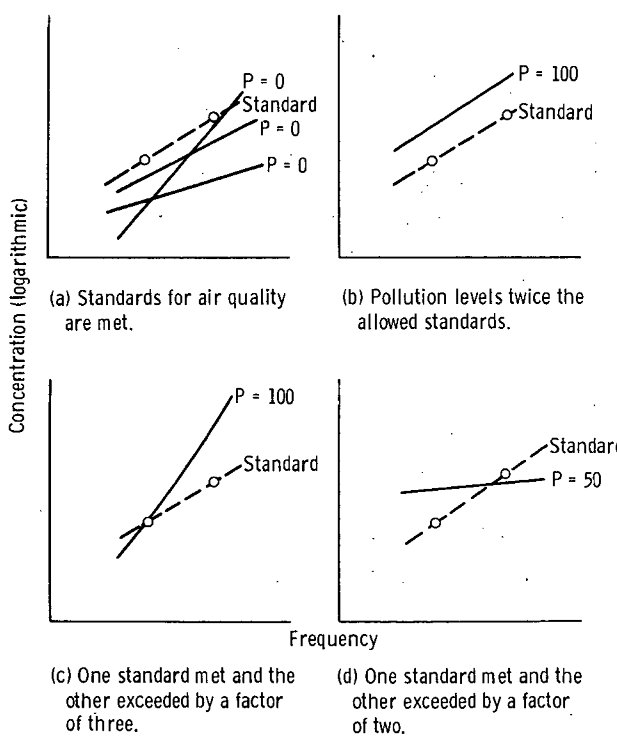


Figure 4. - Examples of Polludex levels.

TABLE VI. - POLLUDEX VALUES FOR 1967 TO 1971

Monitoring station (see fig. 1)	Pollutant	1967	1968	1969	1970	1971
A	Total suspended particulate		408	[a]303	[a]284	296
	Nitrogen dioxide		111	120	114	102
	Sulfur dioxide		[a]201	[a]142	[a]97	70
B	Total suspended particulate	111	103	54	[b]117	82
	Nitrogen dioxide					90
	Sulfur dioxide					5
C	Total suspended particulate	117	[a]144	105	144	167
	Nitrogen dioxide		77	148	134	155
	Sulfur dioxide		103	75	55	49
D	Total suspended particulate	135	135	129	[b]191	(c)
	Nitrogen dioxide		107	119	[b]117	[c]99
	Sulfur dioxide		68	58	[a, b]94	[a, c]64
E	Total suspended particulate	133	[a]159	91	[a]145	[a]109
	Nitrogen dioxide		103	137	117	105
	Sulfur dioxide		85	50	[a]56	26
F	Total suspended particulate	[a]85	104	72	[a]93	89
	Nitrogen dioxide		112	97	115	103
	Sulfur dioxide		[a]40	[a]42	47	27
G	Total suspended particulate		89	[a]66	98	89
	Nitrogen dioxide		101	121	124	103
	Sulfur dioxide		44	7	10	[a]20
H	Total suspended particulate		62	70	106	91
	Nitrogen dioxide		66	125	113	102
	Sulfur dioxide		[a]34	27	34	15
I	Total suspended particulate	[a]255	324	[a]299	321	[a]283
	Nitrogen dioxide		147	153	138	117
	Sulfur dioxide		[a]108	82	[a]70	25
J	Total suspended particulate	203	[a]213	[a]230	[a]207	251
	Nitrogen dioxide			125	155	140
	Sulfur dioxide			99	100	[a]45
K	Total suspended particulate	[a]55	[a]59	43	[b]59	81
	Nitrogen dioxide		62	92	[b]109	83
	Sulfur dioxide		27	11	[b]0	[a]19
L	Total suspended particulate				222	280
	Nitrogen dioxide				120	119
	Sulfur dioxide				141	[a]192
M	Total suspended particulate	61	62	37	70	63
	Nitrogen dioxide		57	68	76	59
	Sulfur dioxide		0	0	9	[a]22
N	Total suspended particulate	205	293	268	[b]436	317
	Nitrogen dioxide				108	[a]127
	Sulfur dioxide				[a]62	[a]105
O	Total suspended particulate	65	71	[a]56	85	116
P	Total suspended particulate	127	146	142	151	145
Q	Total suspended particulate	91	71	60	[a]153	69
R	Total suspended particulate	56	68	62	77	102
S	Total suspended particulate					73
T	Total suspended particulate					380
U	Nitrogen dioxide					[d]129
	Sulfur dioxide					[d]138

[a]The calculation used to obtain this estimate assumed lognormality despite $\sqrt{N}\,D \geq 0.736$.

[b]Sampling site was relocated within same general neighborhood in midyear. It is assumed that for sampling purposes the environmental air was the same at both locations.

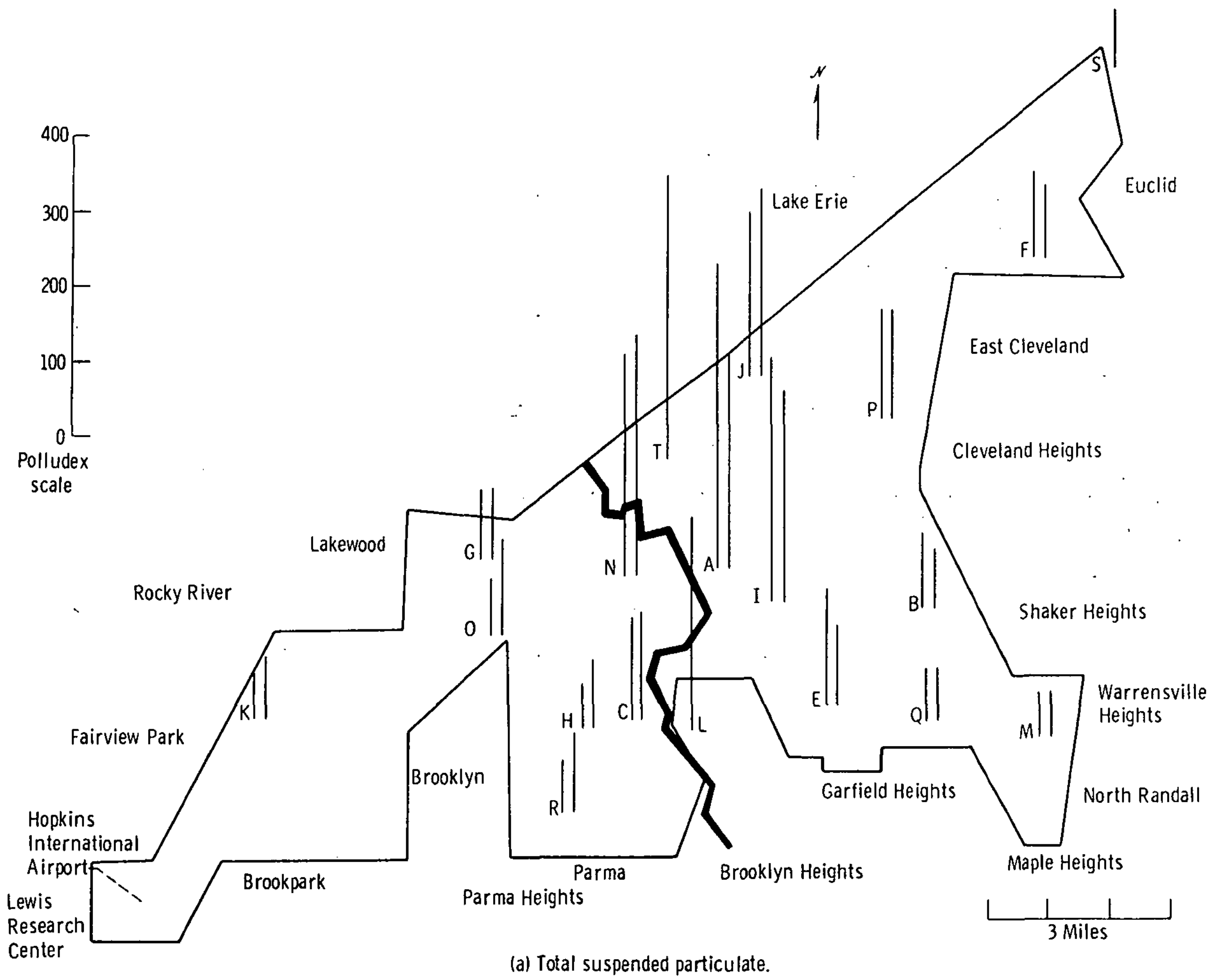
[c]Temporarily discontinued because of construction at sampling site.

[d]Sampling was initiated in the latter part of the year.

Four-Year Trends

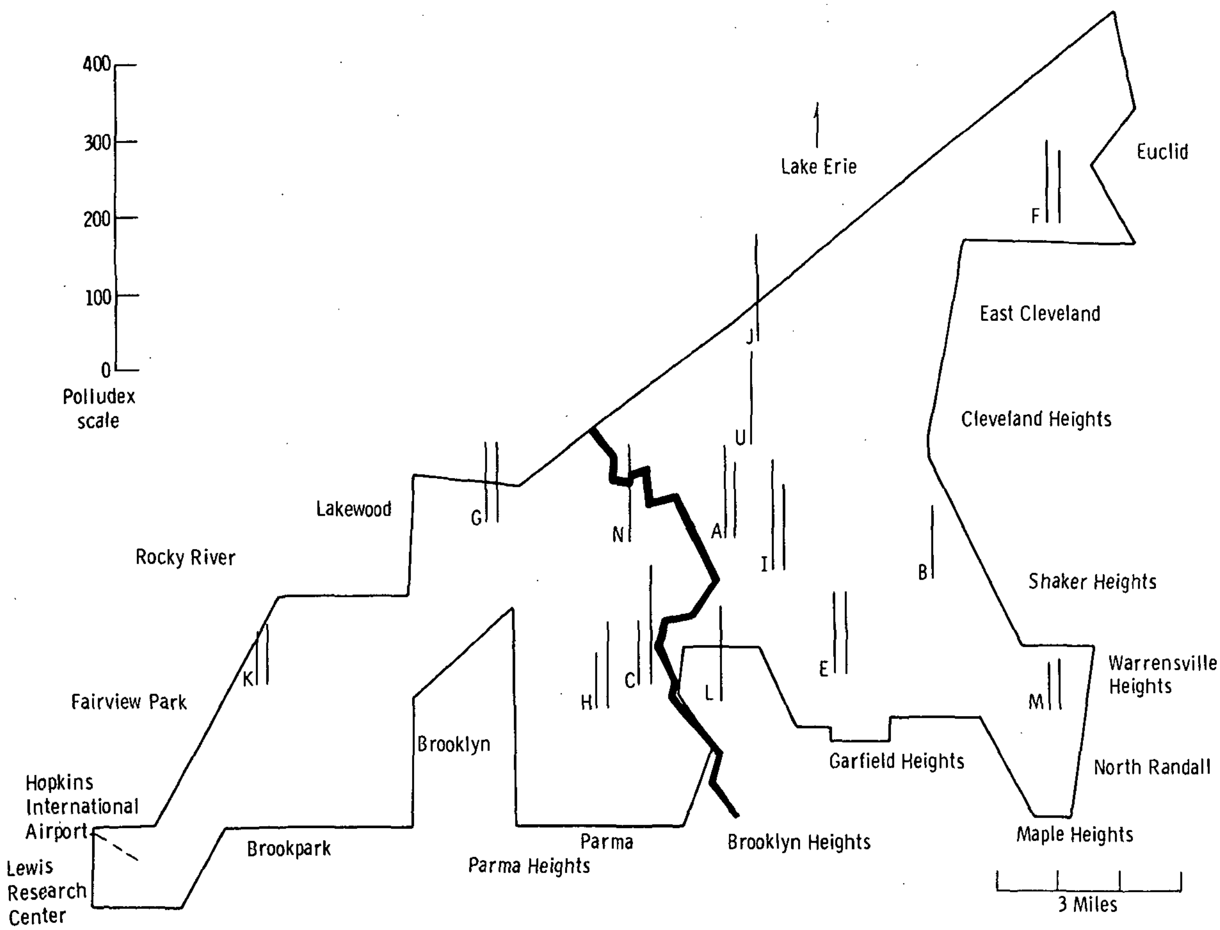
Polludex was evaluated for the APCD data and is listed for all three pollutants in table VI. The State of Ohio standards were used in these calculations.

Where there are adequate data, the 1968 and 1971 values are also presented as bar graphs overprinted on the Cleveland map. The Polludex values for TSP, NO_2, and SO_2 are shown in figures 5(a), (b), and (c), respectively. If there are two bars, the left bar represents 1968 and the right bar 1971. With the exception of site M of figure 5(c), a single bar represents 1971. It is clear that, in general, TSP levels have increased to the west of the Cuyahoga River and decreased to the east. The most pronounced improvements are downwind of the valley (in Cleveland, the winds are predominantly out of the southwest) at sites A, I, and E. The levels of NO_2 show much less variation, ex-



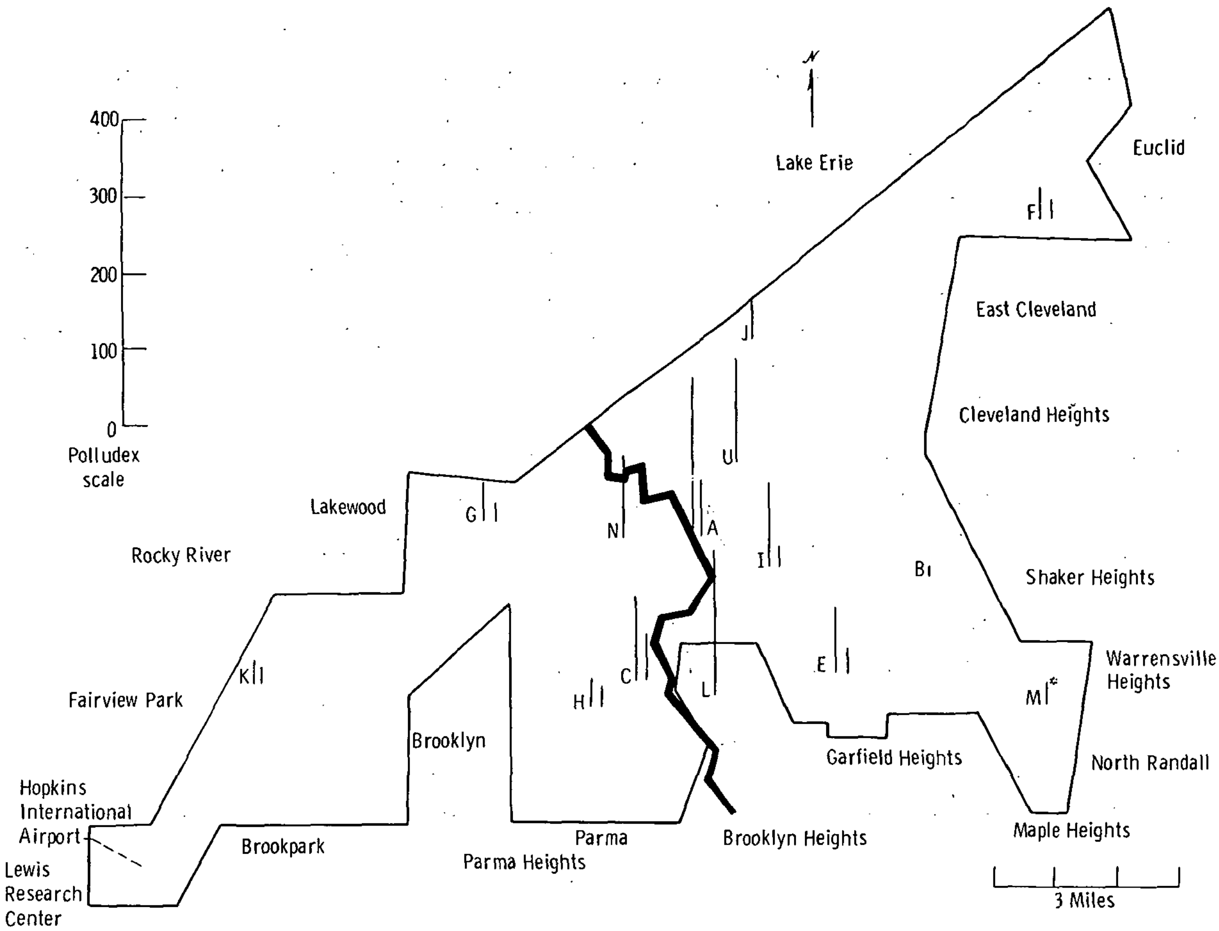
(a) Total suspended particulate.

Figure 5. - Bar graph presentations of Polludex values for the three pollutants at the various monitoring stations. Left bar represents 1968 level of pollution; right bar or a single bar represents 1971 level. Alphabetic coding of monitoring sites corresponds to that of figure 1.



(b) Nitrogen dioxide.

Figure 5. - Continued.



(c) Sulfur dioxide. Value for station M for 1968 was zero.

Figure 5. - Concluded.

cept for the increased levels at sites H and C. With one exception, there has been a significant reduction in the levels of SO_2 throughout the city, with the most pronounced improvements occurring, as with TSP, at sites A, I, and E. Since space heating is fueled primarily by natural gas, this implies a reduction in SO_2 contamination by industrial and power-producing sources. At this time we do not have sufficient information to determine whether the improvements in the valley are due to the general decline in business activity in recent years, the abatement efforts by the industrial community, both of these reasons, or, possibly, neither of these reasons.

CONCLUDING REMARKS

Air-quality data (total suspended particulate, nitrogen dioxide, and sulfur dioxide) for Cleveland, Ohio, for the period of 1968 to 1971 have been collated and subjected to statistical analysis. It is apparent that the data for total suspended particulate and, to a lesser degree, the data for sulfur dioxide and nitrogen dioxide are lognormally distributed. The air-quality standards of the State of Ohio are met only sporadically by sulfur dioxide in isolated residential neighborhoods. The available data indicate that definite improvement in air quality has taken place in the industrial region. Overall, there appears to be a net improvement in air quality, which would be a reflection primarily of the striking reduction in sulfur dioxide levels.

A pollution index has been introduced which directly displays information regarding the degree to which the environmental air conforms to the mandated standards for the environment. As such, it is a useful tool in air-quality monitoring programs.

Lewis Research Center,
National Aeronautics and Space Administration,
Cleveland, Ohio, July 6, 1972,
770-18.

REFERENCES

1. Neustadter, Harold E.; King, Robert B.; Fordyce, J. Stuart; and Burr, John C., Jr.: Air Quality Aerometric Data for the City of Cleveland from 1967 to 1970 for Sulfur Dioxide, Suspended Particulates, and Nitrogen Dioxide. NASA TM X-2496, 1972.
2. Anon.: Federal Register (36FR8186) Apr. 1971.
3. Anon.: State of Ohio, Department of Health, Air Pollution Unit. Regulations AP-3-02, AP-7-01, 1972.
4. Harter, H. Leon: Expected Values of Normal Order Statistics. Biometrika, vol. 48, 1961, pp. 151-165.
5. Noether, Gottfried E.: Elements of Nonparametric Statistics. John Wiley & Sons, Inc., 1967.
6. Lilliefors, Hubert W.: On the Kolmogorov-Smirnov Test for Normality With Mean and Variance Unknown. J. Am. Stat. Assoc., vol. 62, no. 318, June 1967, pp. 399-402.

7. Benarie, M.: Sur la Validitie de la Distribution Logarithmical-normalie des Concentrations de Polluant. Proc. Second Intern, Clean Air Congress, Paper SU-18D, 1970.

8. Mitchell, R. I.: Permanence of the Log-Normal Distribution. J. Opt. Soc. Am., vol. 58, no. 9, Sept. 1968, pp. 1267-1272.

9. Larsen, Ralph I.: A Mathematical Model for Relating Air Quality Measurements to Air Quality Standards. U. S. Rep. AP-89 Environmental Protection Agency, Office of Air Programs, Nov. 1971.

10. Babcock, L. R.: A Combined Pollution Index for Measurement of Total Air Pollution. J. Air Poll. Cont. Assoc., vol. 10, 1970, p. 653.

11. Pikul, Robert: Development of Environmental Indices. Rep. M71-47, Mitre Corp., 1971.

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

— NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION OFFICE

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546